

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
4 November 2004 (04.11.2004)

PCT

(10) International Publication Number  
**WO 2004/095373 A2**

(51) International Patent Classification<sup>7</sup>: G06T 7/00

(21) International Application Number:  
PCT/GB2004/001545

(22) International Filing Date: 8 April 2004 (08.04.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
0308943.0 17 April 2003 (17.04.2003) GB

(71) Applicant, (for all designated States except US): THE  
UNIVERSITY OF DUNDEE [GB/GB]: Nethergate,  
Dundee DD1 4HN (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): ROBERTS, Timo-  
thy [GB/GB]: 16 Castle Street, Tayport, Fife DD6 9AF  
(GB). MCKENNA, Stephen, James [GB/GB]: 5 Newport  
Road, Tayport, Fife DD6 8HP (GB). RICKETTS, Ian,  
Walter [GB/GB]: 78 West Road, Newport on Tay, Fife  
DD6 8HP (GB).

(74) Agent: KENNEDYS PATENT AGENCY LIMITED;  
Floor 5, Queens House, 29 St. Vincent Place, Glasgow G1  
2DT (GB).

(81) Designated States (unless otherwise indicated, for every  
kind of national protection available): AE, AG, AL, AM,  
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,  
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,  
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,  
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,  
MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG,  
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM,  
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM,  
ZW.

(84) Designated States (unless otherwise indicated, for every  
kind of regional protection available): ARIPO (BW, GH,  
GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), Euro-  
pean (AE, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR,  
GB, GR, HU, IE, IT, LU, MC, NE, NL, PL, PT, RO, SE, SI, SK,  
TR), OAPI (BE, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,  
ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished  
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: METHOD AND SYSTEM FOR DETERMINING OBJECT POSE FROM IMAGES

(57) Abstract: A method and system for identifying an object or structured parts of an object in an image. A set of templates are created for each of a number of the parts of the object and the templates are applied to an area of interest in an image where it is hypothesised that an object part is present. The image is analysed to determine the probability that it contains the object part. Thereafter, other templates are applied to other areas of interest in the image to determine the probability that this area of interest belongs to a corresponding object part. The templates are then arranged in a configuration and the likelihood that the configuration represents an object or structured parts of an object is calculated. This is calculated for other configurations and the configuration that is most likely to represent an object or structured part of an object is determined. The method and system can be applied to creating a markerless motion capture system and has other applications in image processing.

BEST AVAILABLE COPY



WO 2004/095373 A2

8/10/05

10/553664

JC09 Rec'd PCT/PTO 17 OCT 2009

WO 2004/095373

PCT/GB2004/001545

1 Method and System for Determining Object Pose from Images

2

3 The present invention relates to a method and system for  
4 determining object pose from images such as still  
5 photographs, films or the like. In particular, the  
6 present invention is designed to allow a user to obtain a  
7 detailed estimation of the pose of a body, particularly a  
8 human body, from real world images with unconstrained  
9 image features.

10

11 In the case of the human body, the task of obtaining pose  
12 information is made difficult because of the large  
13 variation in human appearance. Sources of variation  
14 include the scale, viewpoint, surface texture,  
15 illumination, self-occlusion, object-occlusion, body  
16 structure and clothing shape. In order to deal with  
17 these many complicating factors, it is common, in the  
18 prior art, to use a high level hand built shape model in  
19 which points on this shape model are associated with  
20 image measurements. A score can be computed and a search  
21 performed to find the best solutions to allow the pose of  
22 the body to be determined.

23

1 A second approach identifies parts of the body and then  
2 assembles them into the best configuration. This approach  
3 does not model self-occlusion. Both approaches tend to  
4 rely on a fixed number of parts being parameterised. In  
5 addition, many human pose estimation methods use rigid  
6 geometric primitives such as cones and spheres to model  
7 body parts.

8  
9 Furthermore, existing techniques identify the boundary  
10 between the foreground in which the body part is situated  
11 and the background containing the rest of the scene shown  
12 in the image, by the detection of the edges between these  
13 two features.

14  
15 Where the pose of a body is to be tracked through a  
16 series of images on a frame by frame basis, localised  
17 sampling of the images is used in the full dimensional  
18 pose space. The approach usually requires manual  
19 initialisation and does not recover from significant  
20 tracking errors.

21  
22 It is an object of the present invention to provide an  
23 improved method and system for identifying in an image  
24 the relative positions of parts of a pre-defined object  
25 (object pose) and to use this identification to analyse  
26 images in a number of technological applications areas.

27  
28 In accordance with a first aspect of the present  
29 invention there is provided a method of identifying an  
30 object or structured parts of an object in an image, the  
31 method comprising the steps of:  
32 creating a set of templates, the set containing a  
33 template for each of a number of predetermined object

1 parts and applying said template to an area of interest  
2 in an image where it is hypothesised that an object part  
3 is present;  
4 analysing image pixels in the area of interest to  
5 determine the likelihood that it contains the object  
6 part;  
7 applying other templates from the set of templates to  
8 other areas of interest in the image to determine the  
9 probability that said area of interest belongs to a  
10 corresponding object part and arranging the templates in  
11 a configuration;  
12 calculating the likelihood that the configuration  
13 represents an object or structured parts of an object;  
14 and  
15 calculating other configurations and comparing said  
16 configurations to determine the configuration that is  
17 most likely to represent an object or structured part of  
18 an object.  
19  
20 Preferably, the probability that an area of interest  
21 contains an object part is calculated by calculating a  
22 transformation from the co-ordinates of a pixel in the  
23 area of interest to the template.  
24  
25 Preferably, the step of analysing the area of interest  
26 further comprises identifying the dissimilarity between  
27 foreground and background of the template.  
28  
29 Preferably, the step of analysing the area of interest  
30 further comprises calculating a likelihood ratio based on  
31 a determination of the dissimilarity between foreground  
32 and background features of a transformed template.  
33

1 Preferably, the templates are applied by aligning their  
2 centres, orientations in 2D or 3D and scales to the area  
3 of interest on the image.

4

5 Preferably, the template is a probabilistic region mask  
6 in which values indicate a probability of finding a pixel  
7 corresponding to an object part.

8

9 Optionally, the probabilistic region mask is estimated by  
10 segmentation of training images.

11

12 Optionally, the mask is a binary mask.

13

14 Preferably, the image is an unconstrained scene.

15

16 Preferably, the step of calculating the likelihood that  
17 the configuration represents an object or a structured  
18 part of an object comprises calculating a likelihood  
19 ratio for each object part and calculating the product of  
20 said likelihood ratios.

21

22 Preferably, the step of calculating the likelihood that  
23 the configuration represents an object comprises  
24 determining the spatial relationship of object part  
25 templates.

26

27 Preferably, the step of determining the spatial  
28 relationship of the object part templates comprises  
29 analysing the configuration to identify common boundaries  
30 between pairs of object part templates.

31

32 Optionally, the step of determining the spatial  
33 relationship of the object part templates requires

1 identification of object parts having similar  
2 characteristics and defining these as a sub-set of the  
3 object part templates.

4

5 Preferably, the step of calculating the likelihood that  
6 the configuration represents an object or structured part  
7 of an object comprises calculating a link value for  
8 object parts which are physically connected.

9

10 Preferably, the step of comparing said configurations  
11 comprises iteratively combining the object parts and  
12 predicting larger configurations of body parts.

13

14 Preferably, the object is a human or animal body.

15

16 In accordance with a second aspect of the invention there  
17 is provided a system for identifying an object or  
18 structured parts of an object in an image, the system  
19 comprising:

20 a set of templates, the set containing a template for  
21 each of a number of predetermined object parts  
22 applicable to an area of interest in an image where it is  
23 hypothesised that an object part is present;

24 analysis means for determining the likelihood that the  
25 area of interest contains the object part;

26 configuring means capable of arranging the applied  
27 templates in a configuration;

28 calculating means to calculate the likelihood that the  
29 configuration represents an object or structured parts of  
30 an object for a plurality of configurations; and

31 comparison means to compare configurations so as to  
32 determine the configuration that is most likely to

33 represent an object or structured part of an object.

1

2 Preferably, the system further comprises imaging means  
3 capable of providing an image for analysis.

4

5 More preferably, the imaging means is a stills camera or  
6 a video camera.

7

8 Preferably, the analysis means is provided with means for  
9 identifying the dissimilarity between foreground and  
10 background of the template.

11

12 Preferably, the analysis means calculates the probability  
13 that an area of interest contains an object part by  
14 calculating a transformation from the co-ordinates of a  
15 pixel in the area of interest to the template.

16

17 Preferably, the analysis means calculates a likelihood  
18 ratio based on a determination of the dissimilarity  
19 between foreground and background features of a  
20 transformed template.

21

22 Preferably, the templates are applied by aligning their  
23 centres, orientations (in 2D or 3D) and scales to the  
24 area of interest on the image.

25

26 Preferably, the template is a probabilistic region mask  
27 in which values indicate a probability of finding a pixel  
28 corresponding to an object part.

29

30 Optionally, the probabilistic region mask is estimated by  
31 segmentation of training images.

32

33 Optionally, the mask is a binary mask.



1

2 Preferably, the image is an unconstrained scene.

3

4 Preferably, the calculating means calculates a likelihood  
5 ratio for each object part and calculating the product of  
6 said likelihood ratios.

7

8 Preferably, the likelihood that the configuration  
9 represents an object comprises determining the spatial  
10 relationship of object part templates.

11

12 Preferably, the spatial relationship of the object part  
13 templates is calculated by analysing the configuration to  
14 identify common boundaries between pairs of object part  
15 templates.

16

17 Preferably, the spatial relationship of the object part  
18 templates is determined by identifying object parts  
19 having similar characteristics and defining these as a  
20 sub-set of the object part templates.

21

22 Preferably, the calculating means is capable of  
23 calculating a link value for object parts which are  
24 physically connected.

25

26 Preferably, the calculating means is capable of  
27 iteratively combining the object parts in order to  
28 predict larger configurations of body parts.

29

30 Preferably, the object is a human or animal body.

31

32 In accordance with a third aspect of the present  
33 invention there is provided, a computer program



1 comprising program instructions for causing a computer to  
2 perform the method of the first aspect of the invention.

3

4 Preferably, the computer program is embodied on a  
5 computer readable medium.

6

7 In accordance with a fourth aspect of the present  
8 invention there is provided a carrier having thereon a  
9 computer program comprising computer implementable  
10 instructions for causing a computer to perform the method  
11 of the first aspect of the present invention.

12

13 In accordance with a fifth aspect of the present  
14 invention there is provided a markerless motion capture  
15 system comprising imaging means and a system for  
16 identifying an object or structured parts of an object in  
17 an image of the second aspect of the present invention.

18

19 The present invention will now be described by way of  
20 example only, with reference to the accompanying drawings  
21 in which:

22

23 Figures 1a is a flow diagram showing the operational  
24 steps used in implementing an embodiment of the present  
25 invention and Figure 1b is a detailed flow diagram of the  
26 steps provided in the likelihood module of the present  
27 invention;

28

29 Figures 2a(i) to 2(viii) show a set of templates for a  
30 number of body parts and Figure 2b (i) to (iii) shows a  
31 reduced set of templates;

32

1 Figure 3a shows a lower leg template, Figure 3b shows the  
2 lower leg template on an image and Figure 3c illustrates  
3 the feature distributions of the background and  
4 foreground regions of the image at or near the template;

5  
6 Figure 4a is a graph comparing the probability density of  
7 foreground and background appearance for  $on$  and  $\overline{on}$  ( $\overline{on}$   
8 meaning not on the part) part configurations for a head  
9 template and Figure 4b is a graph of the log of the  
10 resultant likelihood ratio;

11  
12 Figure 5a is a column of typical images from both outdoor  
13 and indoor environments; Figure 5b is a column is a  
14 projection of the positive log likelihood from the masks  
15 or templates and Figure 5c is the projection of positive  
16 log likelihood from the prior art edge based model;

17  
18 Figure 6a is a graph of the spatial variation of the  
19 learnt log likelihood ratios of the present invention and  
20 Figure 6b is a graph of the spatial variation of the  
21 learnt log likelihood ratios of the prior art edge model;

22  
23 Figure 7a is a graph of the probability density for  
24 paired and non-paired configurations and Figure 7b is a  
25 plot of the log of the resulting likelihood ratio;

26  
27 Figure 8a depicts an image of a body in an unconstrained  
28 background and Figure 8b illustrates the projection of  
29 the likelihood ratio for the paired response to a  
30 person's lower right leg image; and

31  
32 Figures 9a to 9d show results from a search for partial  
33 pose configurations.

1

2 The present invention provides a method and system for  
3 identifying an object such as a body in an image. The  
4 technology used to achieve this result is typically a  
5 combination of computer hardware and software.

6

7 Figure 1a shows a flow diagram of an embodiment of the  
8 present invention in which a still photograph of an  
9 unconstrained scene is analysed to identify the position  
10 of an object, in this example, a human body within the  
11 scene.

12

13 Firstly, an image is created 3 using standard  
14 photographic techniques or using digital photography and  
15 the image is transferred 5 into a computer system adapted  
16 to operate the method according to the present invention.  
17 'Configuration prior' is data on the expected  
18 configuration of the body based upon known earlier body  
19 poses or known constraints on body pose such as the basic  
20 stance adopted by a person before taking a golf swing.  
21 This data can be used to assist with the overall analysis  
22 of body pose.

23

24 A configuration hypothesis generator of a known type  
25 creates a configuration 10 created. The likelihood  
26 module 11 creates a score or likelihood 14 which is fed  
27 back to the configuration hypothesis generator 9. Pose  
28 hypotheses are created and a pose output is selected  
29 which is typically the best pose.

30

31 Figure 1b shows the operation of the likelihood generator  
32 in more detail. A geometry analysis module 14 is used to  
33 analyse the geometry of body parts by finding a mask for

1 each part in the configuration and using the  
2 configuration to determine a transformation for each part  
3 from the part's mask to the image and then inverting this  
4 transformation.

5  
6 An appearance builder module 16 is used to analyse the  
7 pixels in an image in the following manner. For every  
8 pixel in the image, the inverse transform is used to find  
9 the corresponding position on each part's mask and the  
10 probability from the mask is used to add the image  
11 features at that image location to the feature  
12 distributions.

13  
14 An appearance evaluation module 18 is used to compare the  
15 foreground and background feature distributions for each  
16 part to get the single part likelihood. The foreground  
17 distributions are compared for each symmetric part to get  
18 the symmetry likelihood. The cues are combined to get the  
19 total likelihood.

20  
21 Details of the manner in which the above embodiment of  
22 the present invention is implemented will now be given  
23 with reference to figures 2 to 9.

24  
25 The shape of each of a number of body parts is modelled  
26 in the following manner. The body part, labelled here by  
27  $i$  ( $i \in 1 \dots N$ ), is represented using a single probabilistic  
28 region template,  $M_i$ , which represents the uncertainty in  
29 the part's shape without attempting to enable shape  
30 instances to be accurately reconstructed. This approach  
31 allows for efficient sampling of the body part shape  
32 where the shape is obscured by a cover if, for example  
33 the subject is wearing loose fitting clothing.

1  
2 The probability that a pixel in the image at position  $(x,$   
3  $y)$  belongs to a hypothesised body part  $i$  is given by  
4  $M_i(T_i(x, y))$  where  $T_i$  is a linear transformation from image  
5 co-ordinates to template or mask co-ordinates determined  
6 by the part's centre,  $(x_c, y_c)$ , image plane rotation,  $\theta$ ,  
7 elongation,  $e$ , and scale,  $s$ . The elongation parameter  
8 alters the aspect ratio of the template and is used to  
9 approximate rotation in depth about one of the part's  
10 axes.

11  
12 The probabilities in the template are estimated from  
13 example shapes in the form of binary masks obtained by  
14 manual segmentation of training images in which the  
15 elongation is maximal (i.e. in which the major axis of  
16 the part is parallel to the image plane). These training  
17 examples are aligned by specifying their centres,  
18 orientations and scales. Un-parameterised pose  
19 variations are marginalised over, allowing a reduction in  
20 the size of the state space. Specifically, rotation  
21 about each limb's major axis is marginalised since these  
22 rotations are difficult to observe. The templates can  
23 also be constrained to be symmetric about their minor  
24 axis.

25  
26 Figures 2a(i) to (viii) show templates with masks for  
27 human body parts. Figure 2a(i) is a mask of a head,  
28 Figure 2a(ii) is a mask of a torso, Figure 2a(iii) is a  
29 mask of an upper arm, Figure 2a(iv) is a mask of a lower  
30 arm, Figure 2a(v) is a mask of a hand, Figure 2a(vi) is a  
31 mask of an upper leg, Figure 2a(vii) is a mask of a lower  
32 leg and Figure 2a(viii) is a mask of a foot.

33

1 In this example, upper and lower arm and leg parts can  
2 reasonably be represented using a single template. This  
3 reduced number of masks greatly improves the sampling  
4 efficiency.

5

6 Figure 2b (i) to (iii) show some learnt probabilistic  
7 region templates. Figure 2b(i) shows a head mask, Figure  
8 2b(ii) shows a torso mask and figure 2b(iii) shows a leg  
9 mask used in this example.

10

11 The uncertain regions in these templates exist because of  
12 (i) 3D shape variation due to change of clothing and  
13 identity of the body, (ii) rotation in depth about the  
14 major axis, and (iii) inaccuracies in the alignment and  
15 manual segmentation of the training images.

16

17 In order to detect the body parts in an image, the  
18 dissimilarity between the appearance of the foreground  
19 and background of a transformed probabilistic region as  
20 illustrated in Fig. 3 is determined. These appearances  
21 are represented as Probability Density Functions (PDFs)  
22 of intensity and chromaticity image features, resulting  
23 in 3D probability distributions.

24

25 In general, local filter responses could also be used to  
26 represent the appearance. Since texture can often result  
27 in multi-modal distributions, each PDF is encoded as a  
28 histogram (marginalised over position). For scenes in  
29 which the body parts appear small, semi-parametric  
30 density estimation methods such as Gaussian mixture  
31 models can be used.

32

1 The foreground appearance histogram for part  $i$ , denoted  
 2 here by  $F_i$ , is formed by adding image features from the  
 3 part's supporting region proportional to  $M_i(T_i(x, y))$ .  
 4 Similarly, the adjacent background appearance  
 5 distribution,  $B_i$ , is estimated by adding features  
 6 proportional to  $1 - M_i(T_i(x, y))$ .  
 7  
 8 The foreground appearance will be less similar to the  
 9 background appearance for configurations that are correct  
 10 (denoted by  $on$ ) than incorrect (denoted by  $\overline{on}$ ).  
 11 Therefore, a PDF of the Bhattacharya measure (for  
 12 measuring the divergence of the probability density  
 13 functions) given by Equation (1) is learnt for  $on$  and  $\overline{on}$   
 14 configurations.  
 15  
 16 The  $on$  distribution is estimated from data obtained by  
 17 specifying the transformation parameters to align the  
 18 probabilistic region template to be on parts that are  
 19 neither occluded nor overlapping. The  $\overline{on}$  distribution is  
 20 estimated by generating random alignments elsewhere in  
 21 sample images of outdoor and indoor scenes.  
 22  
 23 The  $on$  PDF can be adequately represented by a Guassian  
 24 distribution. Equation (2) defines  $SINGLE_i$  as the ratio  
 25 of the  $on$  and  $\overline{on}$  distributions. This is used to score a  
 26 single body part configuration and is plotted in Fig. 3.  
 27

$$I(F_i, B_i) = \sum_f \sqrt{F(f) \times B(f)} \quad (1)$$

$$SINGLE_i = \frac{p(I(F_i, B_i)|on)}{p(I(F_i, B_i)|\overline{on})} \quad (2)$$



1 Figure 4a is a graph comparing the probability density of  
2 foreground and background appearance for  $on$  and  $\overline{on}$  part  
3 configurations for a head template and Figure 4b is a  
4 graph of the log of the resultant likelihood ratio.

5 It is clear from Figure 3a that the probability density  
6 distributions for the  $on$  and  $\overline{on}$  distributions are well  
7 separated.

8  
9 The present invention also provides enhanced  
10 discrimination of body parts by defining adjoining and  
11 non-adjoining regions.

12  
13 Detection of single body parts, can be improved by  
14 distinguishing positions where the background appearance  
15 is most likely to differ from the foreground appearance.  
16 For example, due to the structure of clothing, when  
17 detecting an upper arm, adjoining background areas around  
18 the shoulder joint are often similar to the foreground  
19 appearance. The histogram model proposed thus far, which  
20 marginalises appearance over position, does not use this  
21 information optimally.

22  
23 To enhance discrimination, two separate adjacent  
24 background histograms are constructed, one for adjoining  
25 regions and another for non-adjoining regions. In the  
26 model, it is expected that the non-adjoining region  
27 appearance will be less similar to the foreground  
28 appearance than the adjoining region appearance.

29  
30 The adjoining and non-adjoining regions can be specified  
31 manually during training by defining a hard threshold.  
32 Alternatively, a probabilistic approach, where the

1 regions are estimated by marginalising over the relative  
2 pose between adjoining parts to get a low dimensional  
3 model could be used.

4

5 The use of information from adjoining regions is  
6 particularly useful where bottom-up identification of  
7 body parts is required.

8

9 Figures 5a to 5c show a set of images (Figure 5a) which  
10 have been analysed for part detection purposes using the  
11 present invention (Figure 5b) and by using a prior art  
12 method (Figure 4c). Figure 5a is a column of typical  
13 images from both outdoor and indoor environments, Figure  
14 5b is a column is a projection of the positive log  
15 likelihood from the masks or templates showing the  
16 maximum likelihood of the presence of body parts and  
17 Figure 5c is the projection of positive log likelihood  
18 from the prior art edge based model.

19

20 The column Fig. 5b shows the projection of the likelihood  
21 ratio computed using Equation (2) onto typical images  
22 containing significant background information or clutter.  
23 The top image of Figure 5b shows the response for a head  
24 while the other two images show the response of a  
25 vertically-orientated limb filter.

26

27 It can be seen that the technique of the present  
28 invention is highly discriminatory, producing relatively  
29 few false maxima in comparison with the prior art system.  
30 Although images were acquired using various cameras, some  
31 with noisy colour signals, system parameters were fixed  
32 for all test images.

33

1 In order to provide a comparison with an alternative  
2 method, the responses obtained by comparing the  
3 hypothesised part boundaries with edge responses were  
4 computed. These are shown in Fig. 5c. Orientations of  
5 significant edge responses for foreground and background  
6 configurations were learned (using derivatives of the  
7 probabilistic region template), treated as independent  
8 and normalised for scale. Contrast normalisation was not  
9 used. Other formulations (e.g. averaging) proved to be  
10 weaker on the scenes under consideration. The responses  
11 using this method are clearly less discriminatory.

12

13 Figures 6a and 6b compare the spatial variation of the  
14 Log of Learnt likelihood ratios of the present invention  
15 and the prior art edge-based likelihood system for a  
16 head. In both Figures 6a and 6b, the correct position is  
17 centred and indicated by the vertical line 25. The  
18 horizontal bar 27 in both Figures 6a and 6b corresponds  
19 to a likelihood ratio of more than 1 which is the measure  
20 of whether an object is more likely to be a head than  
21 not. As can be seen from comparing Figures 6a and 6b,  
22 Figure 6b has a large number of positions where the  
23 likelihood is greater than 1, whereas only a single  
24 instance of this occurs in Figure 6a.

25

26 The edge response, whilst indicative of the correct  
27 position of body parts, has significant false positive  
28 likelihood ratios. The part likelihood calculation used  
29 in the present invention is more expensive to compute,  
30 however, it is far more discriminatory and as a result,  
31 fewer samples are needed when performing pose search,  
32 leading to an overall computational performance benefit.  
33 Furthermore, the collected foreground histograms can be

1 useful for other likelihood measurements as described  
2 below.

3  
4 Since any single body part likelihood will probably  
5 result in false positives, the present invention provides  
6 for the encoding of higher order relationships between  
7 body parts to improve discrimination. This is  
8 accomplished by encoding an expectation of structure in  
9 the foreground appearance and the spatial relationship of  
10 body parts.

11  
12 Configurations containing more than one body part can be  
13 represented using an extension of the probabilistic  
14 region approach described above. In order to account for  
15 self-occlusion, the pose space is represented by a depth  
16 ordered set,  $V$ , of probabilistic regions with parts  
17 sharing a common scale parameter,  $s$ . When taken  
18 together, the templates determine the probability that a  
19 particular image feature belongs to a particular part's  
20 foreground or background. More specifically, the  
21 probability that an image feature at position  $(x,y)$   
22 belongs to the foreground appearance of part  $i$  is given  
23 by  $M_i(T_i(x,y)) \times \prod_j (1 - M_j(T_j(x,y)))$  where  $j$  labels closer,  
24 instantiated parts.

25  
26 Therefore, a list of paired body parts is specified and  
27 the background appearance histogram is constructed from  
28 features weighted by  $\prod_k (1 - M_k(T_k(x,y)))$  where  $k$  labels all  
29 instantiated parts other than  $i$  and those paired with  $i$ .

30  
31 Thus, a single image feature can contribute to the  
32 foreground and adjacent background appearance of several  
33 parts. When insufficient data is available to estimate

1 either the foreground or the adjacent background  
 2 histogram (as determined using an area threshold) the  
 3 corresponding likelihood ratio is set to one.

4

5 In order to define constraints between parts, a link is  
 6 introduced between parts  $i$  and  $j$  if and only if they are  
 7 physically connected neighbours. Each part has a set of  
 8 control points that link it to its neighbours. A link  
 9 has an associated value  $LINK_{i,j}$  given by:

$$LINK_{i,j} = \begin{cases} 1 & \text{if } \delta_{i,j}/s < \Delta_{i,j} \\ e^{(\delta_{i,j}/s - \Delta_{i,j})/\sigma} & \text{otherwise} \end{cases} \quad (3)$$

10

11 where  $\delta_{i,j}$  is the image distance between the control  
 12 points of the pair,  $\Delta_{i,j}$  is the maximum un-penalised  
 13 distance and  $\sigma$  relates to the strength of penalisation.  
 14 If the neighbouring parts do not link directly, because  
 15 intervening parts are not instantiated, the un-penalised  
 16 distance is found by summing the un-penalised distances  
 17 over the complete chain. This can be interpreted as  
 18 being analogous to a force between parts equivalent to a  
 19 telescopic rod with a spring on each end.

20

21 A simplifying feature of the system is that certain pairs  
 22 of body parts can be expected to have a similar  
 23 foreground appearance to one another. For example, a  
 24 person's upper left arm will nearly always have a similar  
 25 colour and texture to the person's upper right arm. In  
 26 the system of the present invention, the limbs are paired  
 27 with their opposing parts. To encode this knowledge, a  
 28 PDF of the divergence measure (computed using Equation  
 29 (1)) between the foreground appearance histograms of  
 30 paired parts and non-paired parts is learnt.

Equation (4) shows the resulting likelihood ratio and Figures 7a and 7b describe this ratio graphically. Figure 7a shows a plot of the learnt PDFs of the foreground appearance similarity for paired and non-paired configurations. The log of the resulting likelihood ratio is shown in Figure 7b. The higher probability of similarity is found for the paired configurations.

Figure 8 shows a typical image projection of this ratio and shows the technique to be highly discriminatory. It limits possible configurations if one limb can be found reliably and helps reduce the likelihood of incorrect large assemblies.

$$PAIR_{i,j} = \frac{p(I(F_i, F_j) | on_i, on_j)}{p(I(F_i, F_j) | \overline{on_i}, \overline{on_j})} \quad (4)$$

Learning the likelihood ratios allows a principled fusion of the various cues and principled comparison of the various hypothesised configurations. The individual likelihood ratios are combined by treating the individual likelihood ratios as being independent of one another. The overall likelihood ratio is given by Equation (5). This rewards correct higher dimensional configurations over correct lower dimensional ones.

$$R = \prod_{i \in v} SINGLE_i \times \prod_{i,j \in v} PAIR_{i,j} \times \prod_{l,j \in v} LINK_{l,j} \quad (5)$$

As is apparent from the above equation, the present invention enables different hypothesised configurations to have differing numbers of parts and yet allows a

1 comparison to be made between them in order to decide  
2 which (partial) configuration to infer given the image  
3 evidence.

4  
5 The parts in the inferred configuration may not be  
6 directly physically connected (e.g. the inferred  
7 configuration might consist of a lower leg, an arm and a  
8 head in a given scene either because the other parts are  
9 occluded or their boundaries are not readily apparent  
10 from the image).

11  
12 An example of a sampling scheme useable with the present  
13 invention is described as follows.

14  
15 A coarse regular scan of the image for the head and limbs  
16 is made and these results are then locally optimised.  
17 Part configurations are sampled from the resulting  
18 distribution and combined to form larger configurations  
19 which are then optimised for a fixed period of time in  
20 the full dimensional pose space.

21  
22 Due to the flexibility of the parameterisation, a set of  
23 optimization methods such as genetic style combination,  
24 prediction, local search, re-ordering and re-labelling  
25 can be combined using a scheduling algorithm and a shared  
26 sample population to achieve rapid, robust, global, high  
27 dimensional pose estimation.

28  
29 Fig. 9 shows results of searching for partial pose  
30 configurations. The areas enclosed by the white lines 31,  
31 33, 35, 37, 39, 41, 43, 45, 47 and 49 identify these pose  
32 configurations. Although inter-part links are not  
33 visualised in this example, these results represent



1 estimates of *pose configurations* with inter-part  
2 connectivity as opposed to independently detected parts.  
3 The scale of the model was fixed and the elongation  
4 parameter was constrained to be above 0.7.

5

6 The system of the present invention described above  
7 allows detailed, efficient estimation of human pose from  
8 real-world images.

9

10 The invention provides (i) a formulation that allows the  
11 representation and comparison of partial (lower  
12 dimensional) solutions and models other object occlusion  
13 and (ii) a highly discriminatory learnt likelihood based  
14 upon probabilistic regions that allows efficient body  
15 part detection.

16

17 The likelihood depends only on there being differences  
18 between a hypothesised part's foreground appearance and  
19 adjacent background appearance. The present invention  
20 does not make use of scene-specific background models and  
21 is, as such, general and applicable to unconstrained  
22 scenes.

23

24 The system can be used to locate and estimate the pose of  
25 a person in a single monocular image. In other examples,  
26 the present invention can be used during tracking of the  
27 person in a sequence of images by combining it with a  
28 temporal pose prior propagated from other images in the  
29 sequence. In this example, it allows tracking of the  
30 body parts to reinitialise after partial or full  
31 occlusion or after tracking of certain body parts fails  
32 temporarily for some other reason.

33

1 In a further embodiment, the present invention can be  
2 used in a multi-camera system to estimate the person's  
3 pose from several views captured simultaneously.

4

5 Many other applications follow from this ability to  
6 identify a body or structured parts of a body in an image  
7 (body pose information). In one embodiment of the  
8 present invention, the body pose information determined  
9 can be used as control inputs to drive a computer game or  
10 some other motion-driven or gesture-driven human-computer  
11 interface.

12

13 In another embodiment of the present invention, the body  
14 pose information can be used to control computer  
15 graphics, for example, an avatar.

16

17 In another embodiment of the present invention,  
18 information on the body pose of a person obtained from an  
19 image can be used in the context of an art installation  
20 or a museum installation to enable the installation to  
21 respond interactively to the person's body movements.

22

23 In another embodiment of the present invention, the  
24 detection and pose estimation of people in video images  
25 in particular can be used as part of automated monitoring  
26 and surveillance applications such as security or care of  
27 the elderly.

28

29 In another embodiment of the present invention, the  
30 system could be used as part of a markerless motion-  
31 capture system for use in animation for entertainment and  
32 gait analysis. In particular, it could be used to  
33 analyse golf swings or other sports actions. The system

1 could also be used to analyse image/video archives or as  
2 part of an image indexing system.

3  
4 Some of the features of the invention can be modified or  
5 replaced by alternatives. For example, the use of  
6 histograms could be replaced by some other method of  
7 estimating a frequency distribution (e.g. mixture models,  
8 Parzen windows) or feature representation. Different  
9 methods for comparing feature representations could be  
10 used (e.g. chi-squared, histogram intersection).

11  
12 The part detectors could use other features (e.g.  
13 responses of local filters such as gradient filters,  
14 Gaussian derivatives or Gabor functions).

15  
16 The parts could be parameterised to model perspective  
17 projection. The search over configurations could  
18 incorporate any number of the widely known methods for  
19 high-dimensional search instead of or in combination with  
20 the methods mentioned above.

21  
22 The population-based search could use any number of  
23 heuristics to help bootstrap the search (e.g. background  
24 subtraction, skin colour or other prior appearance  
25 models, change/motion detection).

26  
27 The system presented here is novel in several respects.  
28 The formulation allows differing numbers of parts to be  
29 parameterised and allows poses of differing  
30 dimensionality to be compared in a principled manner  
31 based upon learnt likelihood ratios. In contrast with  
32 current approaches, this allows a part based search in  
33 the presence of self-occlusion. Furthermore, it provides

1 a principled automatic approach to other object  
2 occlusion. View based probabilistic models of body part  
3 shapes are learnt that represent intra and inter person  
4 variability (in contrast to rigid geometric primitives).

5  
6 The probabilistic region template for each part is  
7 transformed into the image using the configuration  
8 hypothesis. The probabilistic region is also used to  
9 collect the appearance distributions for the part's  
10 foreground and adjacent background. Likelihood ratios  
11 for single parts are learnt from the dissimilarity of the  
12 foreground and adjacent background appearance  
13 distributions. This technique does not use restrictive  
14 foreground/background specific modelling.

15  
16 The present invention describes better discrimination of  
17 body parts in real world images than contour to edge  
18 matching techniques. Furthermore, the use of likelihoods  
19 is less sparse and noisy, making coarse sampling and  
20 local search more effective.

21  
22 Improvements and modifications may be incorporated herein  
23 without deviating from the scope of the invention.

24

1   Claims

2

3   1.   A method of identifying an object or structured  
4   parts of an object in an image, the method comprising the  
5   steps of:

6   creating a set of templates, the set containing a  
7   template for each of a number of predetermined object  
8   parts and applying said template to an area of interest  
9   in an image where it is hypothesised that an object part  
10   is present;

11   analysing image pixels in the area of interest to  
12   determine the probability that it contains the object  
13   part;

14   applying other templates from the set of templates to  
15   other areas of interest in the image to determine the  
16   probability that said area of interest belongs to a  
17   corresponding object part and arranging the templates in  
18   a configuration;

19   calculating the likelihood that the configuration  
20   represents an object or structured parts of an object;  
21   and

22   calculating other configurations and comparing said  
23   configurations to determine the configuration that is  
24   most likely to represent an object or structured part of  
25   an object.

26

27   2.   A method as claimed in Claim 1 wherein, the  
28   probability that an area of interest contains an object  
29   part is calculated by calculating a transformation from  
30   the co-ordinates of a pixel in the area of interest to  
31   the template.

32

1 3. A method as claimed in Claim 1 or Claim 2 wherein,  
2 analysing the area of interest further comprises  
3 identifying the dissimilarity between foreground and  
4 background of a transformed probabilistic region.  
5

6 4. A method as claimed in any preceding claim wherein,  
7 analysing the area of interest further comprises  
8 calculating a likelihood ratio based on a determination  
9 of the dissimilarity between foreground and background  
10 features of a transformed template.  
11

12 5. A method as claimed in any preceding claim wherein,  
13 the templates are applied by aligning their centres,  
14 orientations in 2D or 3D and scales to the area of  
15 interest on the image.  
16

17 6. A method as claimed in any preceding Claim wherein  
18 the template is a probabilistic region mask in which  
19 values indicate a probability of finding a pixel  
20 corresponding to an object part.  
21

22 7. A method as claimed in any preceding claim wherein,  
23 the probabilistic region mask is estimated by  
24 segmentation of training images.  
25

26 8. A method as claimed in any preceding claim wherein,  
27 the image is an unconstrained scene.  
28

29 9. A method as claimed in any preceding claim wherein,  
30 the step of calculating the likelihood that the  
31 configuration represents an object or a structured part  
32 of an object comprises calculating a likelihood ratio for

1 each object part and calculating the product of said  
2 likelihood ratios.

3

4 10. A method as claimed in any preceding claim wherein,  
5 the step of calculating the likelihood that the  
6 configuration represents an object comprises determining  
7 the spatial relationship of object part templates.

8

9 11. A method as claimed in Claim 10 wherein the step of  
10 determining the spatial relationship of the object part  
11 templates comprises analysing the configuration to  
12 identify common boundaries between pairs of object part  
13 templates.

14

15 12. A method as claimed in Claim 11 wherein the step of  
16 determining the spatial relationship of the object part  
17 templates requires identification of object parts having  
18 similar characteristics and defining these as a sub-set  
19 of the object part templates.

20

21 13. A method as claimed in any preceding claim, wherein  
22 the step of calculating the likelihood that the  
23 configuration represents an object or structured part of  
24 an object comprises calculating a link value for object  
25 parts which are physically connected.

26

27 14. A method as claimed in any preceding claim wherein  
28 the step of comparing said configurations comprises  
29 iteratively combining the object parts and predicting  
30 larger configurations of body parts.

31

32 15. A method as claimed in any preceding claim wherein  
33 the object is a human or animal body.



- 1  
2 16. A system for identifying an object or structured  
3 parts of an object in an image, the system comprising:  
4 a set of templates, the set containing a template for  
5 each of a number of predetermined object parts  
6 applicable to an area of interest in an image where it is  
7 hypothesised that an object part is present;  
8 analysis means for determining the probability that the  
9 area of interest contains the object part;  
10 configuring means capable of arranging the applied  
11 templates in a configuration;  
12 calculating means to calculate the likelihood that the  
13 configuration represents an object or structured parts of  
14 an object for a plurality of configurations; and  
15 comparison means to compare configurations so as to  
16 determine the configuration that is most likely to  
17 represent an object or structured part of an object.  
18
- 19 17. A system as claimed in Claim 16 wherein, the system  
20 further comprises imaging means capable of providing an  
21 image for analysis.  
22
- 23 18. A system as claimed in claim 17 wherein the imaging  
24 means is a stills camera or a video camera.  
25
- 26 19. A system as claimed in Claims 16 to 18 wherein, the  
27 analysis means is provided with means for identifying the  
28 dissimilarity between foreground and background of a  
29 transformed probabilistic region.  
30
- 31 20. A system as claimed in Claims 16 to 19 wherein, the  
32 analysis means calculates the probability that an area of  
33 interest contains an object part by calculating a

1 transformation from the co-ordinates of a pixel in the  
2 area of interest to the template.

3

4 21. A method as claimed in any of Claims 16 to 20  
5 wherein, the analysis means calculates a likelihood ratio  
6 based on a determination of the dissimilarity between  
7 foreground and background features of a transformed  
8 template.

9

10 22. A system as claimed in Claims 16 to 21 wherein, the  
11 templates are applied by aligning their centres,  
12 orientations (in 2D or 3D) and scales to the area of  
13 interest on the image.

14

15 23. A system as claimed in any of Claims 16 to 22  
16 wherein the template is a probabilistic region mask in  
17 which values indicate a probability of finding a pixel  
18 corresponding to the body part.

19

20 24. A system as claimed in any of Claims 16 to 22  
21 wherein, the probabilistic region mask is estimated by  
22 segmentation of training images.

23

24 25. A system as claimed in Claims 16 to 24 wherein, the  
25 image is an unconstrained scene.

26

27 26. A system as claimed in Claims 16 to 25 wherein, the  
28 calculating means calculates a likelihood ratio for each  
29 object part and calculating the product of said  
30 likelihood ratios.

31

32 27. A system as claimed in Claim 26 wherein, the  
33 likelihood that the configuration represents an object

1 comprises determining the spatial relationship of object  
2 part templates.

3

4 28. A system as claimed in Claim 27 wherein the spatial  
5 relationship of the object part templates is calculated  
6 by analysing the configuration to identify common  
7 boundaries between pairs of object part templates.

8

9 29. A system as claimed in Claim 28 wherein the spatial  
10 relationship of the object part templates is determined  
11 by identifying object parts having similar  
12 characteristics and defining these as a sub-set of the  
13 object part templates.

14

15 30. A system as claimed in any preceding claim, wherein  
16 the calculating means is capable of calculating a link  
17 value for object parts which are physically connected.

18

19 32. A system as claimed in any of claims 16 to 31  
20 wherein the calculating means is capable of iteratively  
21 combining the object parts in order to predict larger  
22 configurations of body parts.

23

24 33. A method as claimed in Claims 16 to 32 wherein the  
25 object is a human or animal body.

26

27 34. A computer program comprising program instructions  
28 for causing a computer to perform the method of any of  
29 Claims 1 to 15.

30

31 35. A computer program as claimed in claim 34 wherein  
32 the computer program is embodied on a computer readable  
33 medium.

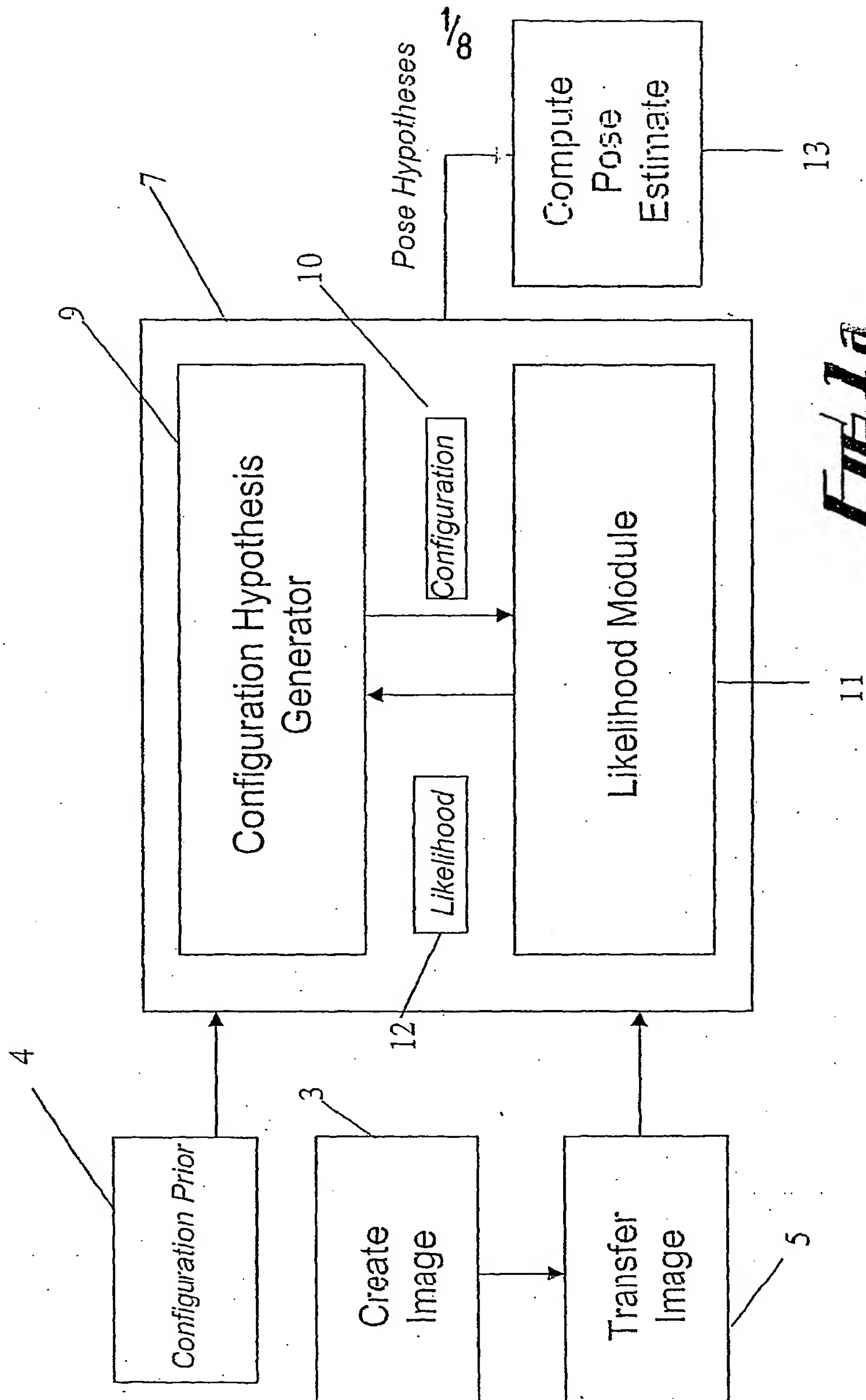
1 36. A carrier having thereon a computer program  
2 comprising computer implementable instructions for  
3 causing a computer to perform the method of any of claims  
4 1 to 15.

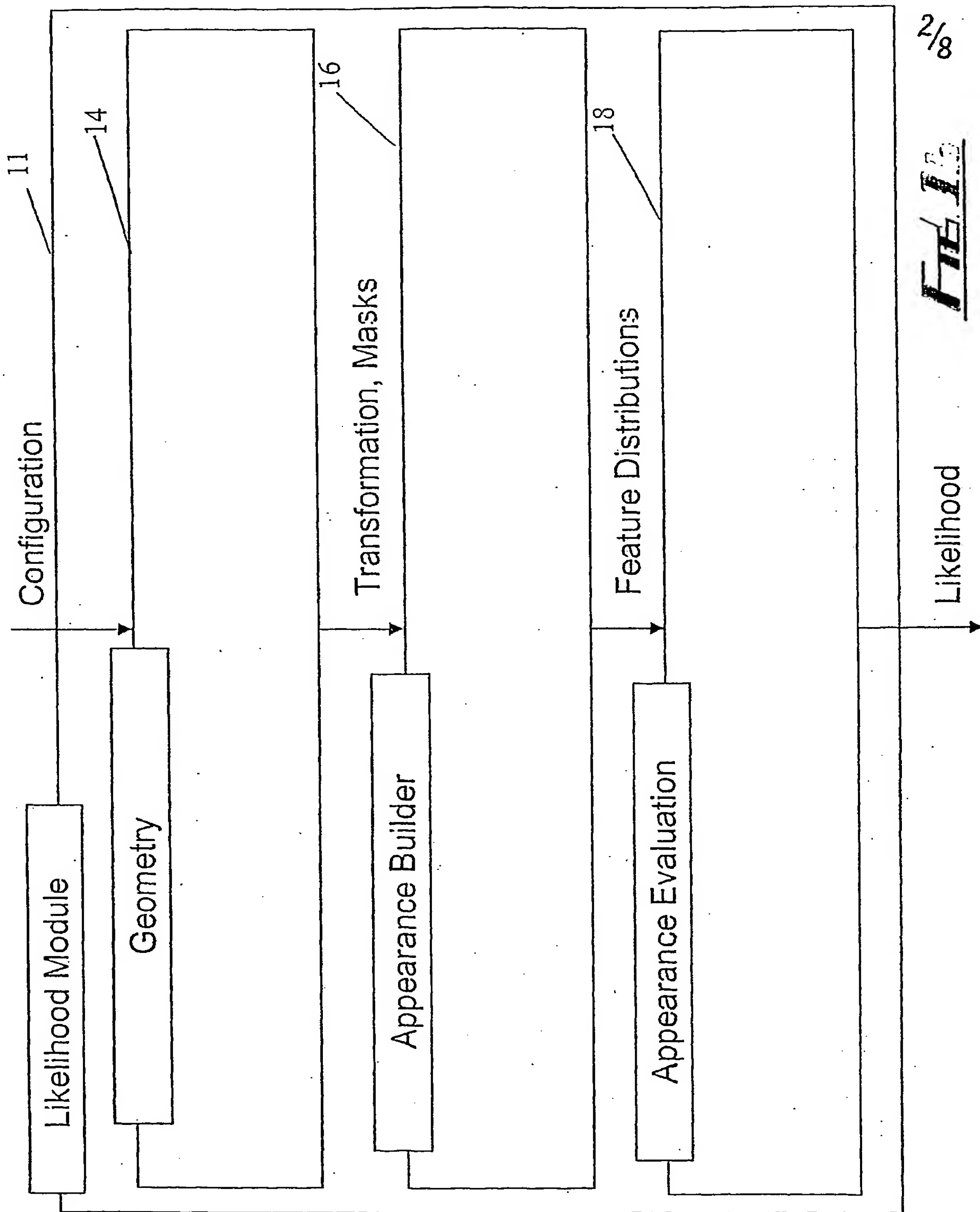
5

6 37. A markerless motion capture system comprising  
7 imaging means and a system for identifying an object or  
8 structured parts of an object in an image as claimed in  
9 any of Claims 16 to 33.

10

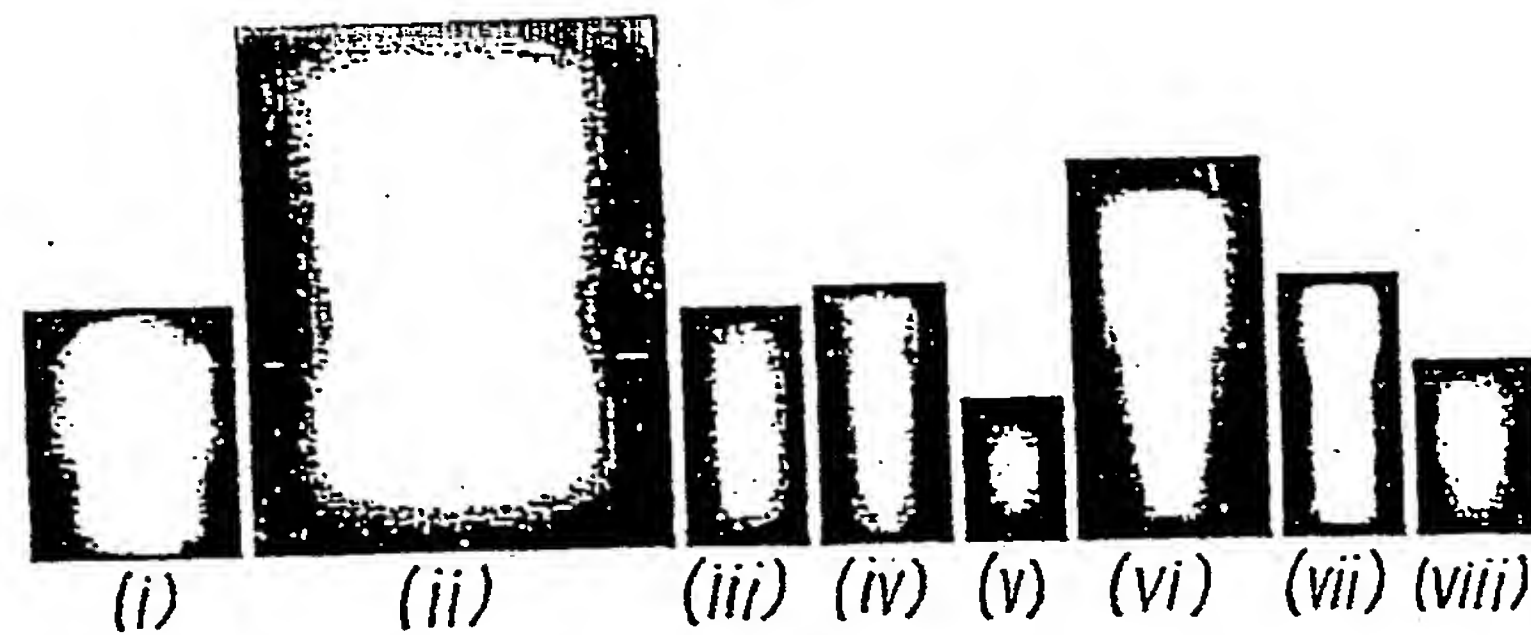
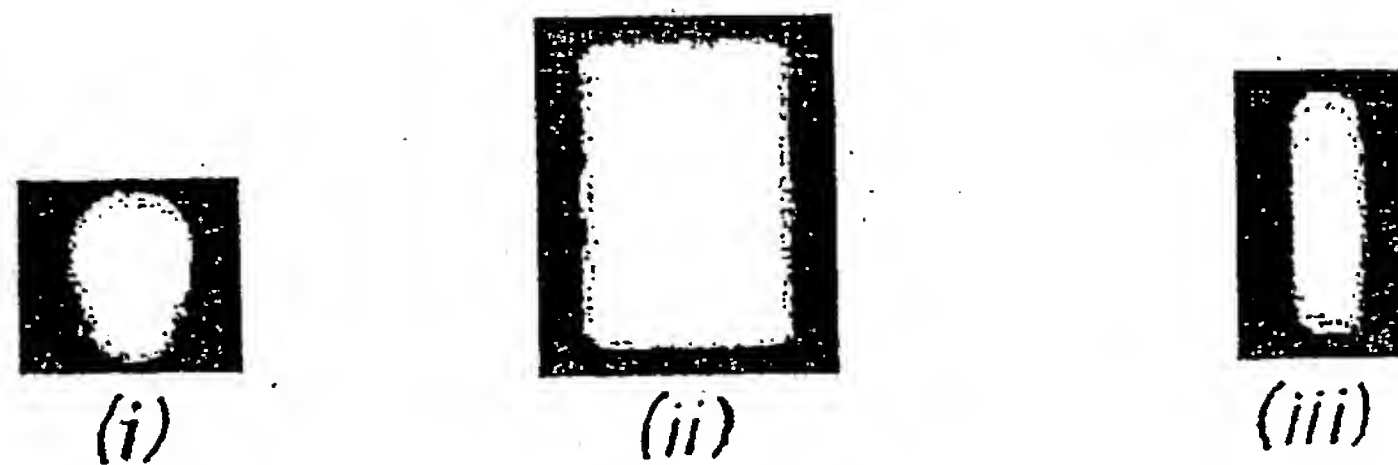
11



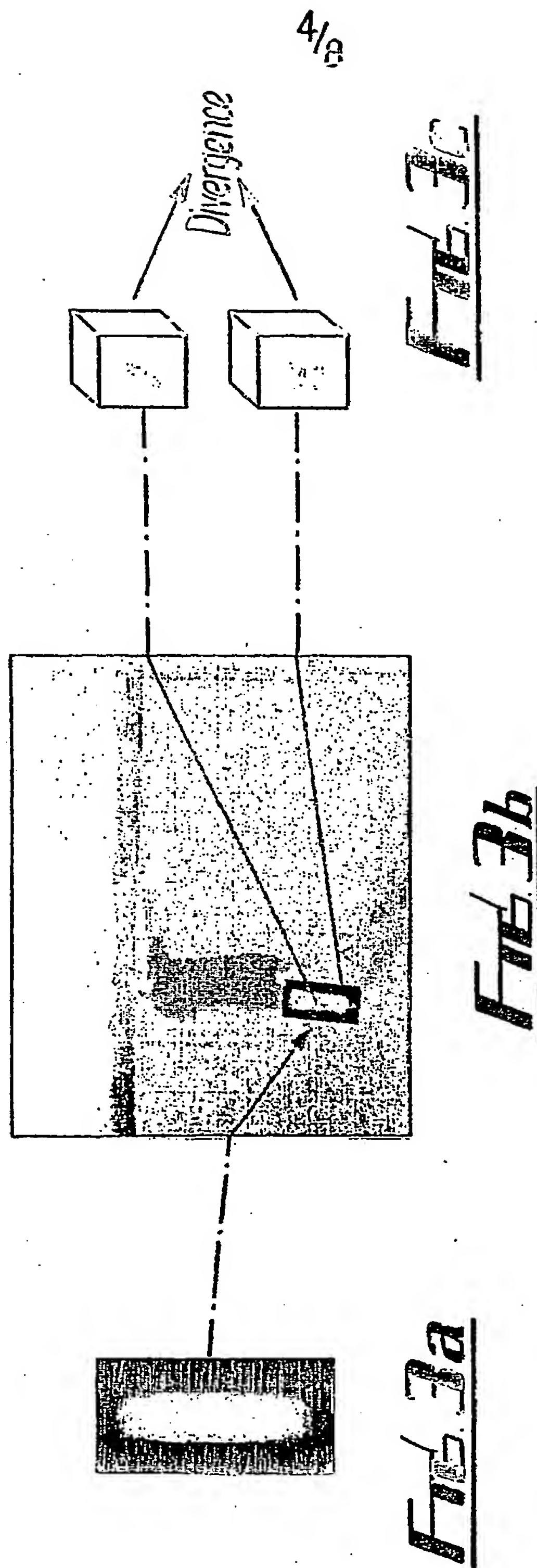


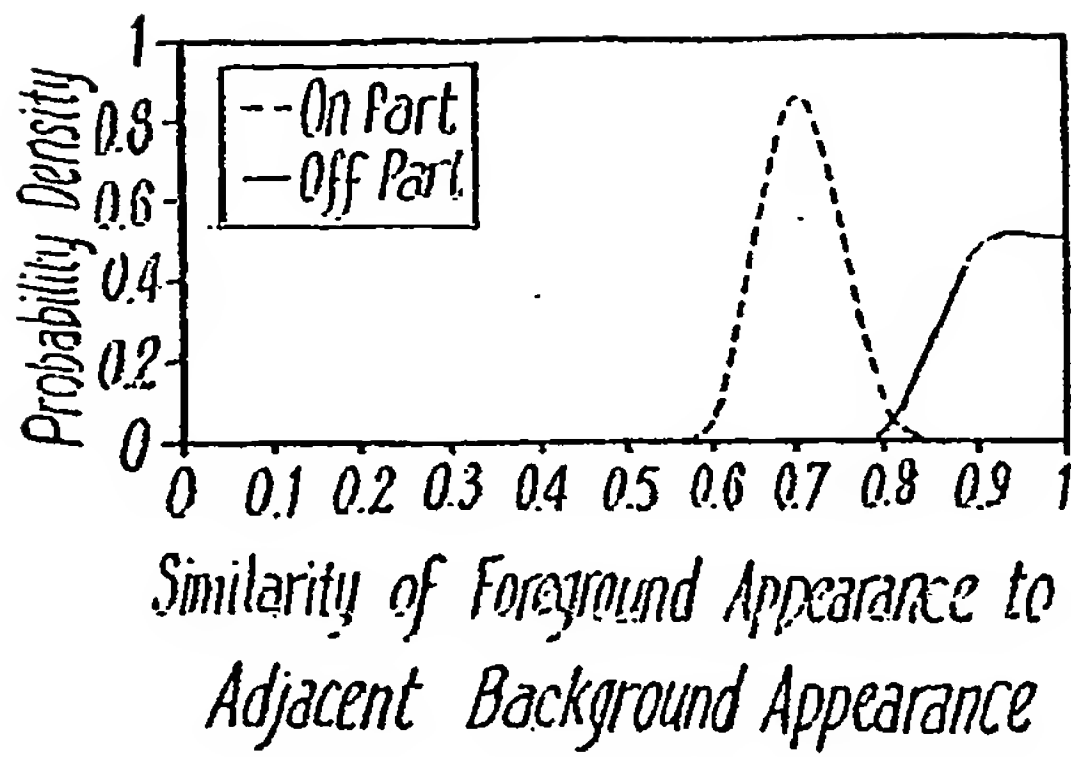
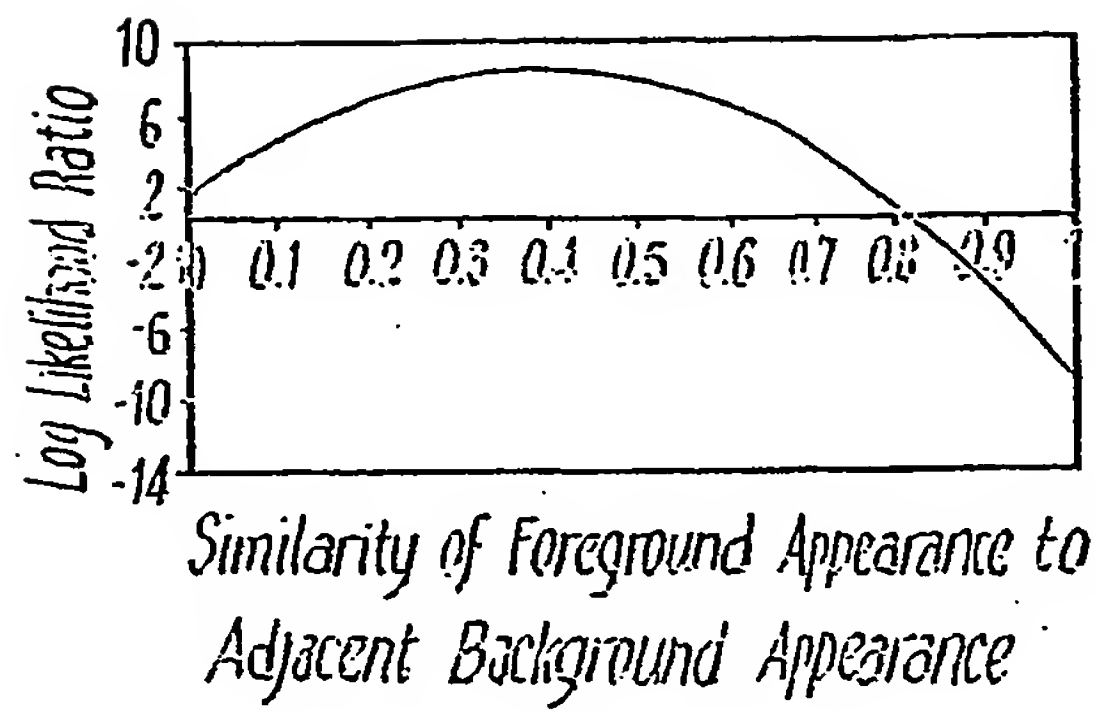
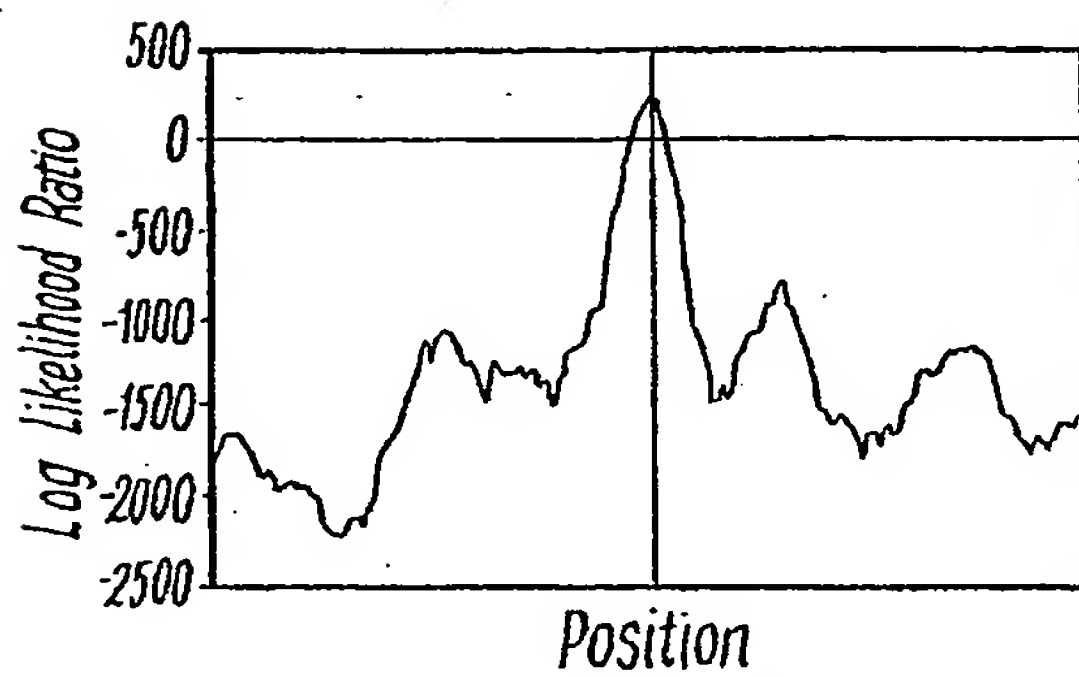
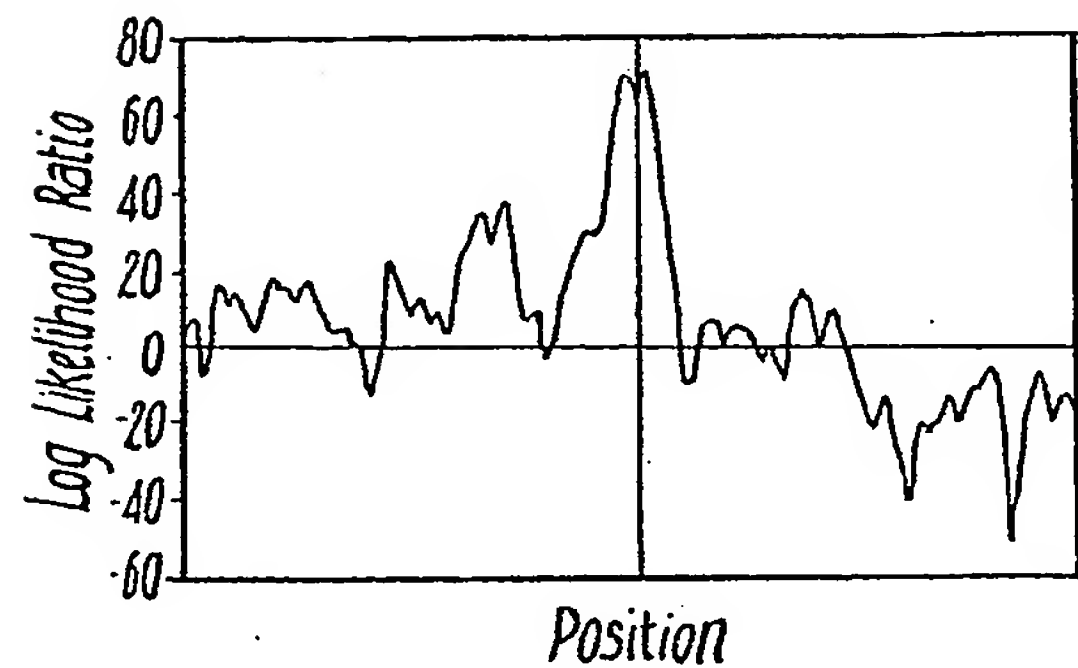
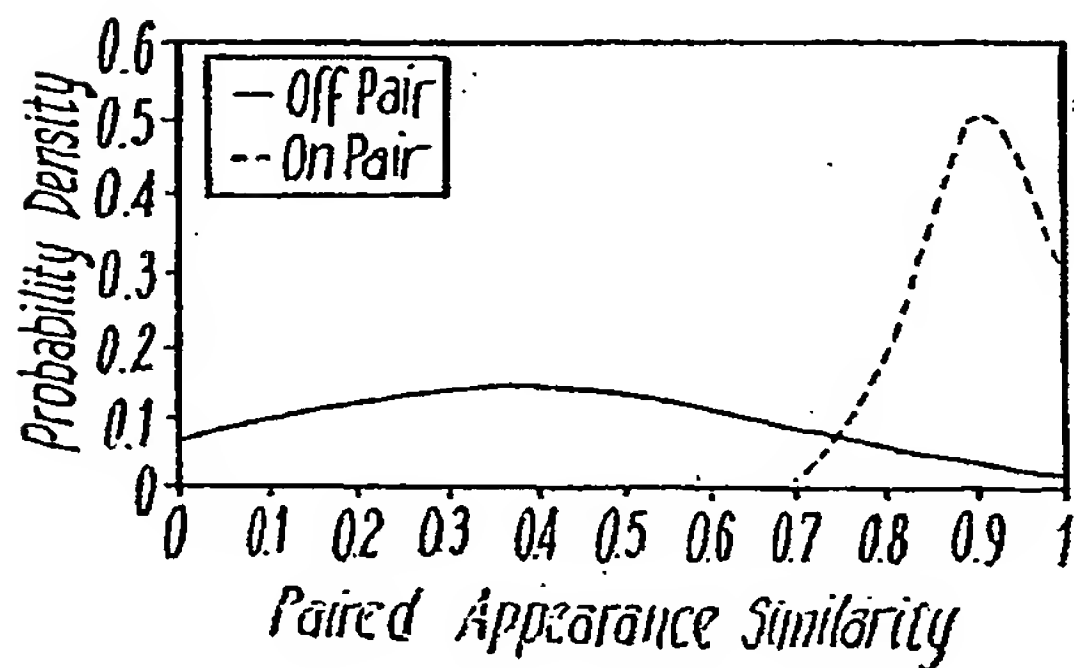
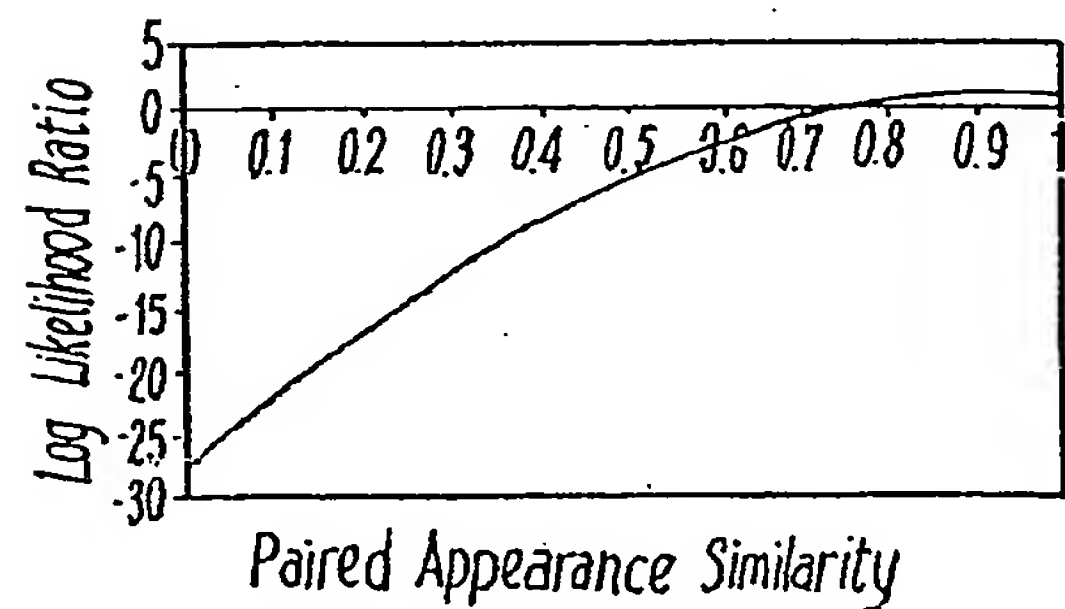
2/8

**Fig. 13**

$\frac{3}{8}$ **FIG. 2a****FIG. 2b**



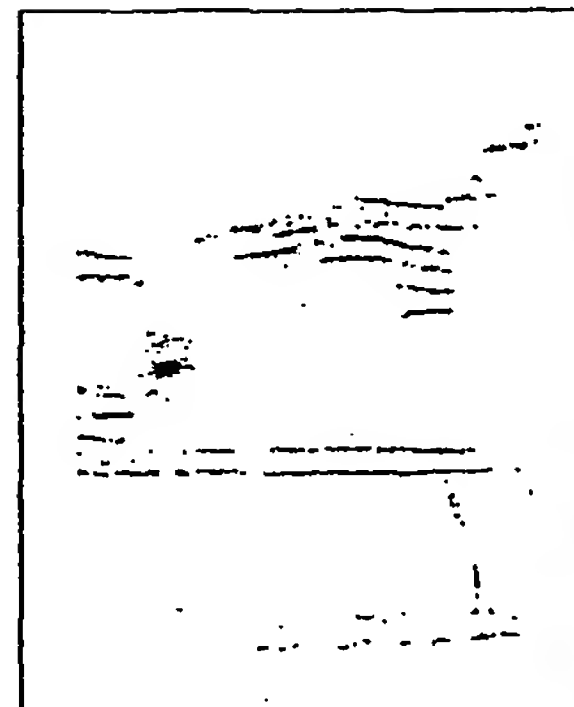
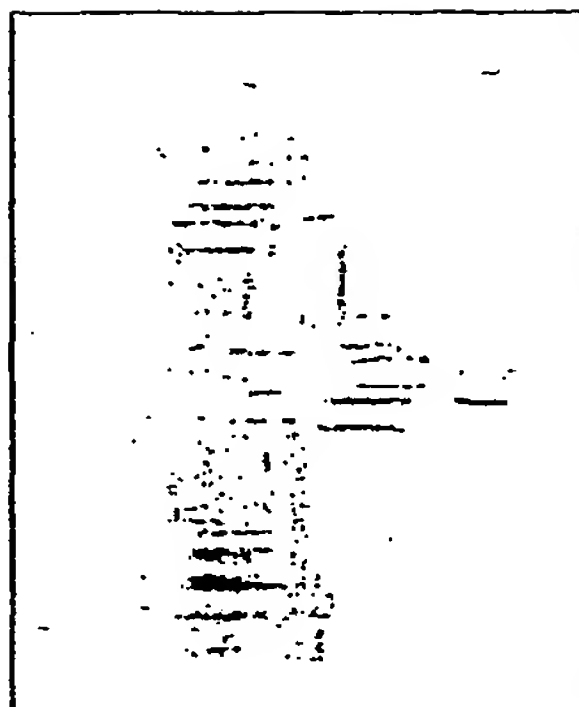
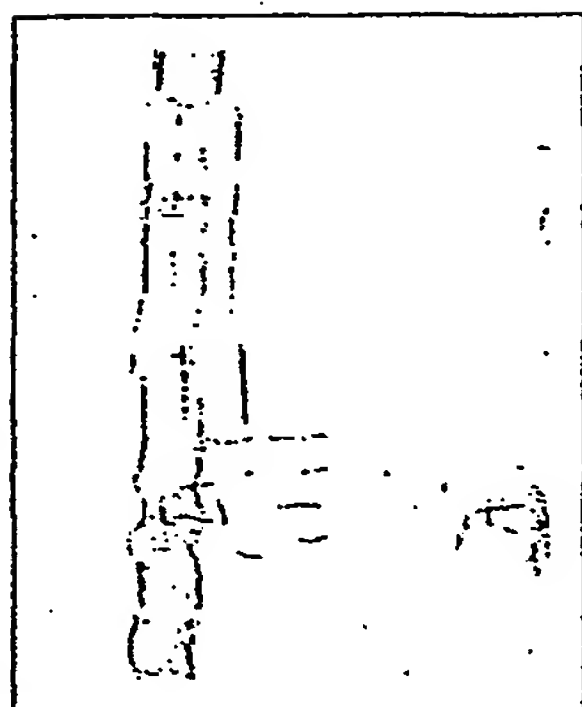


**FIG. 4a****FIG. 4b****FIG. 6a****FIG. 6b****FIG. 7a****FIG. 7b**

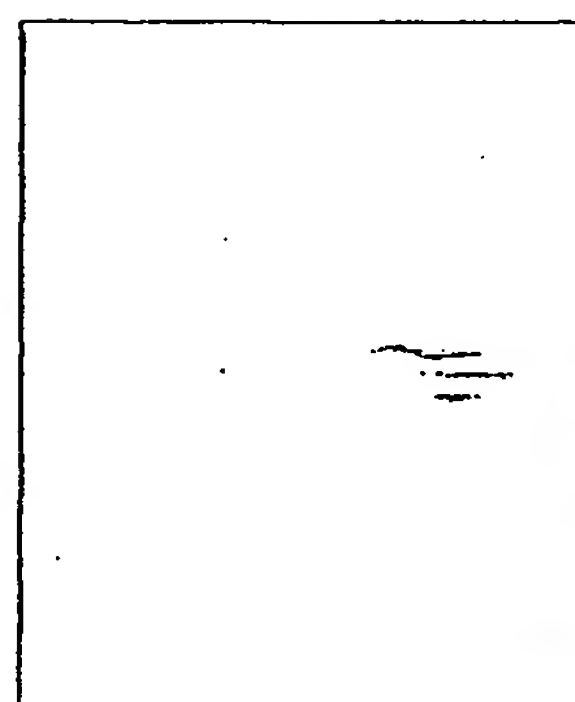
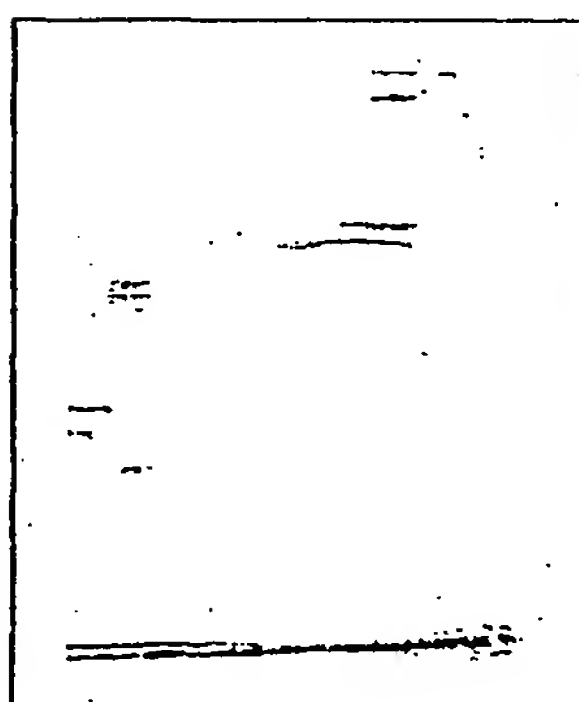
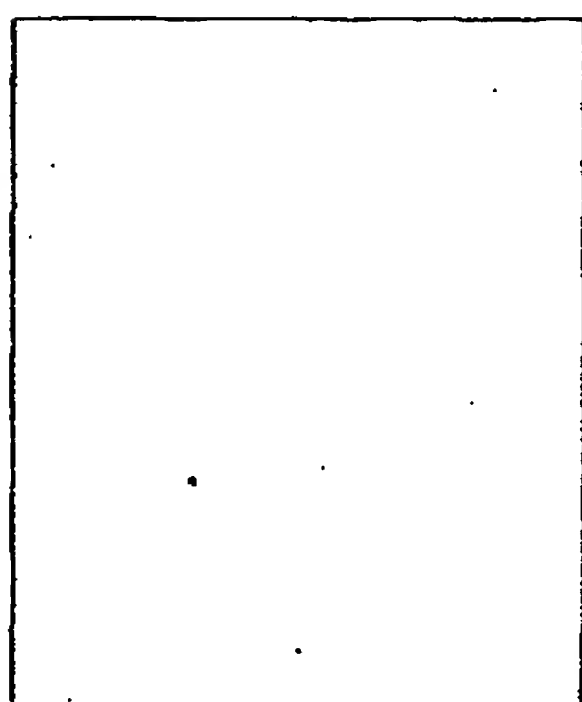
WO 2004/095373

PCT/GB2004/001545

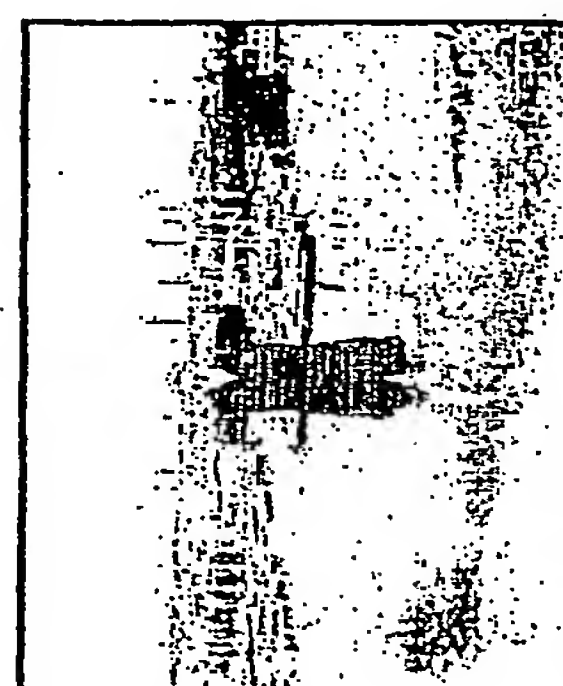
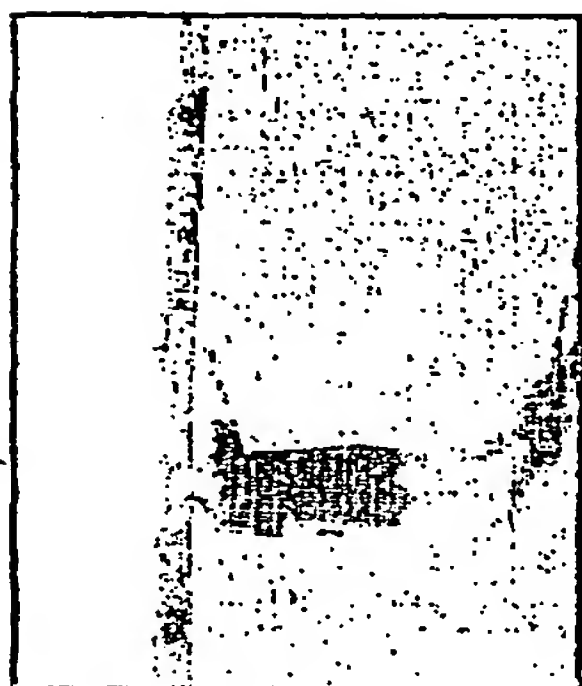
6/8



**Fig. 5c**



**Fig. 5b**

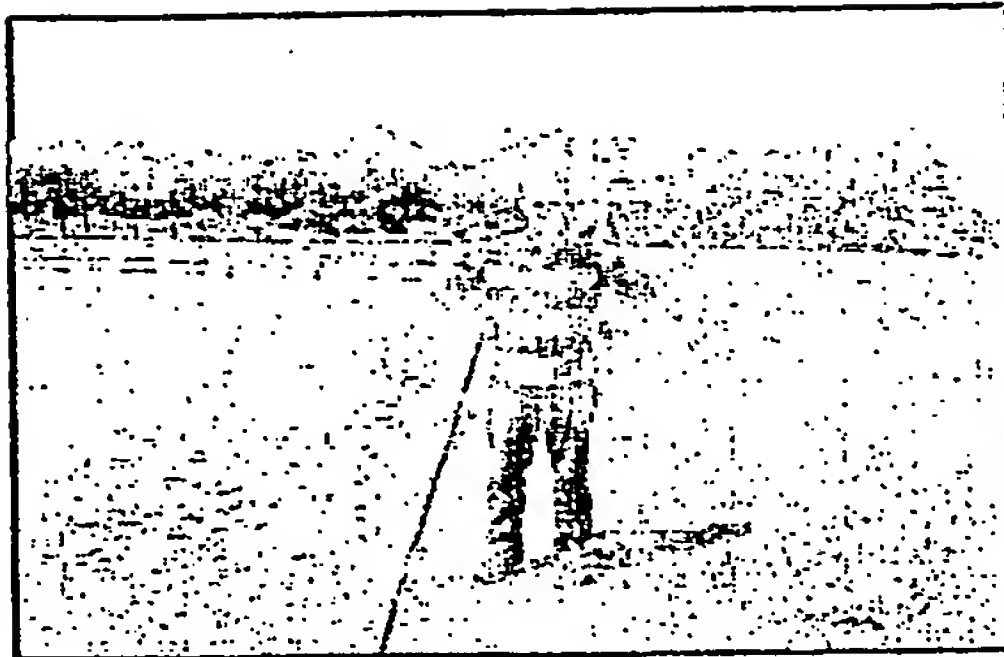


**Fig. 5a**

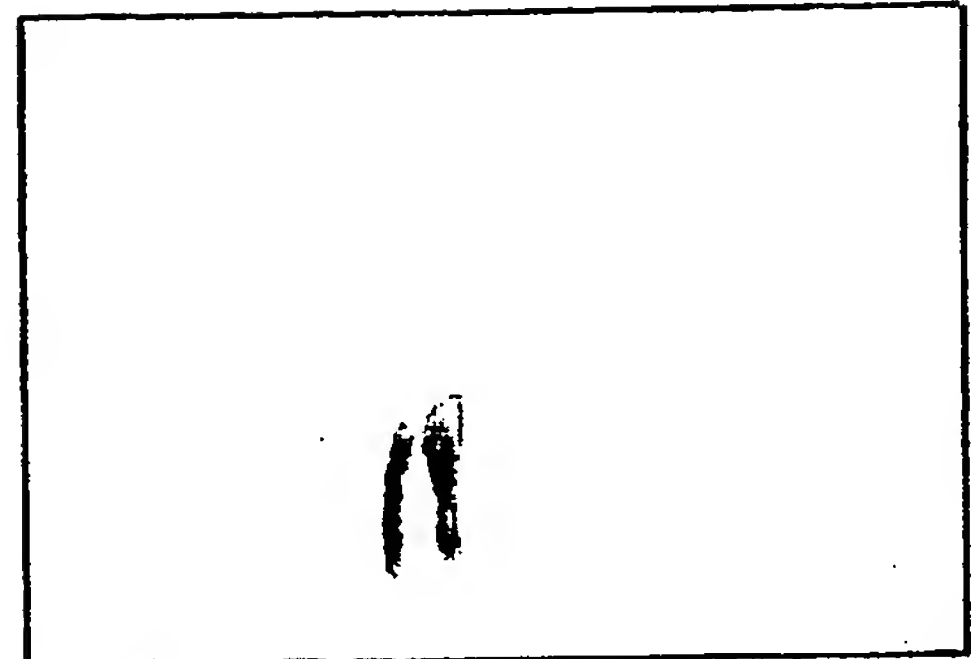
WO 2004/095373

PCT/GB2004/001545

7/8



*Fig. 8a*



*Fig. 8b*

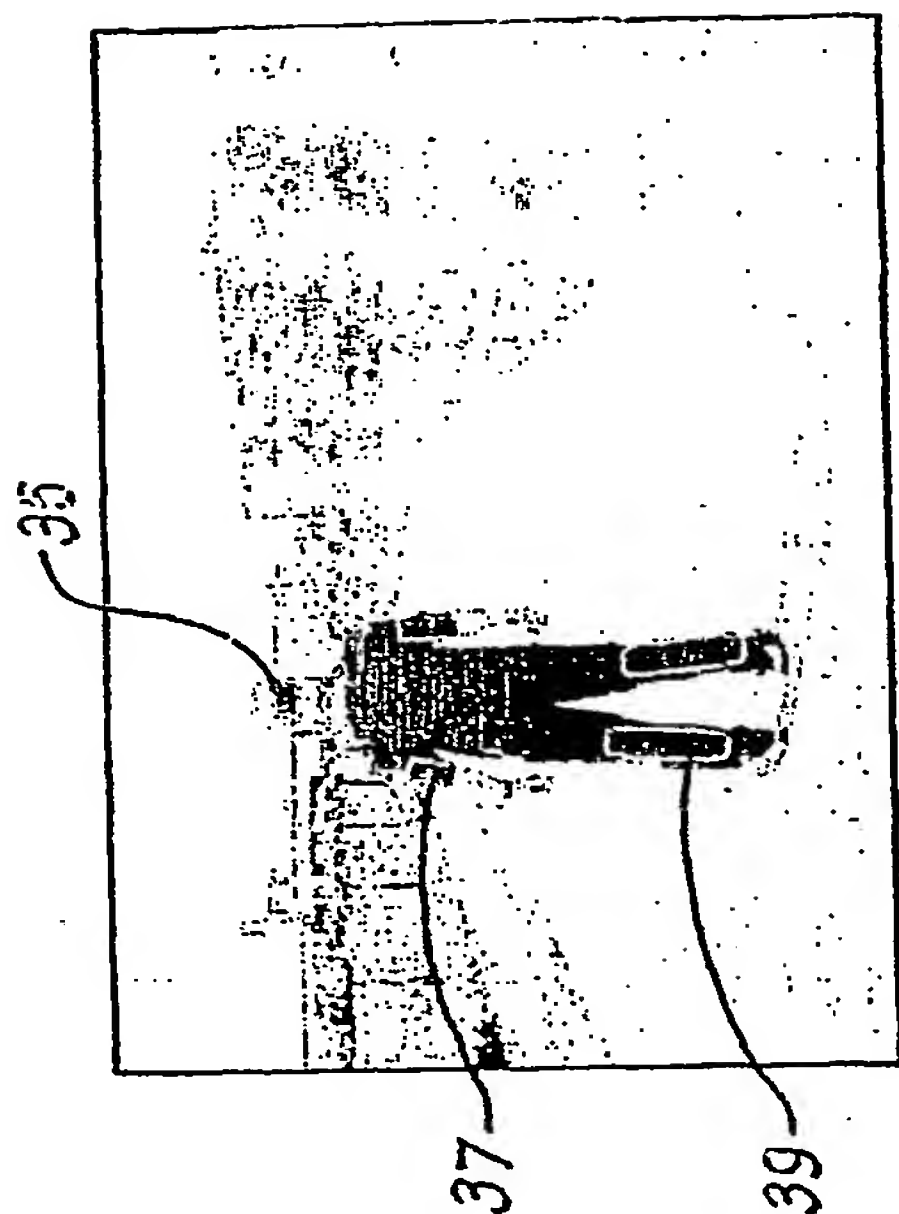


Fig. 9b

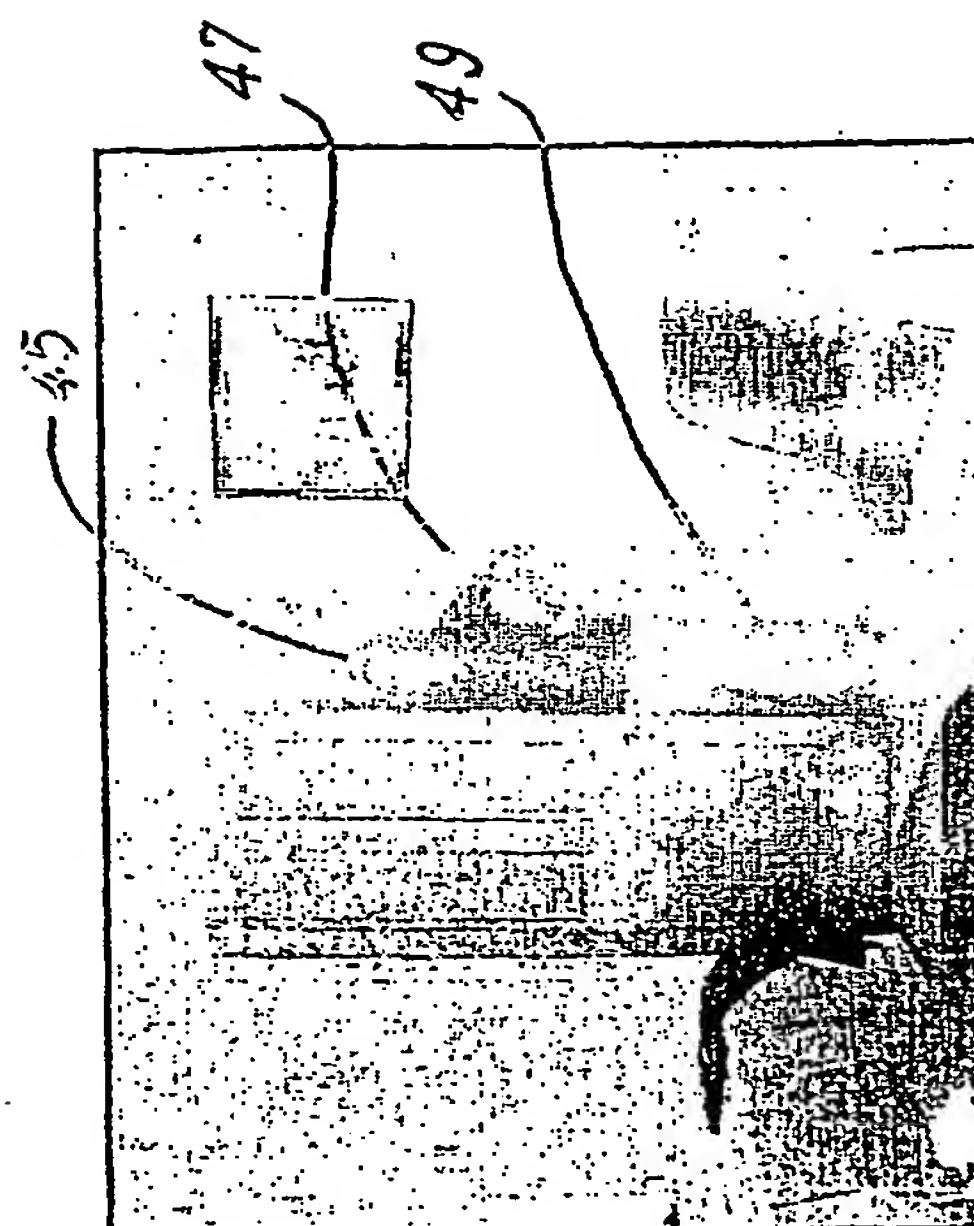


Fig. 9d

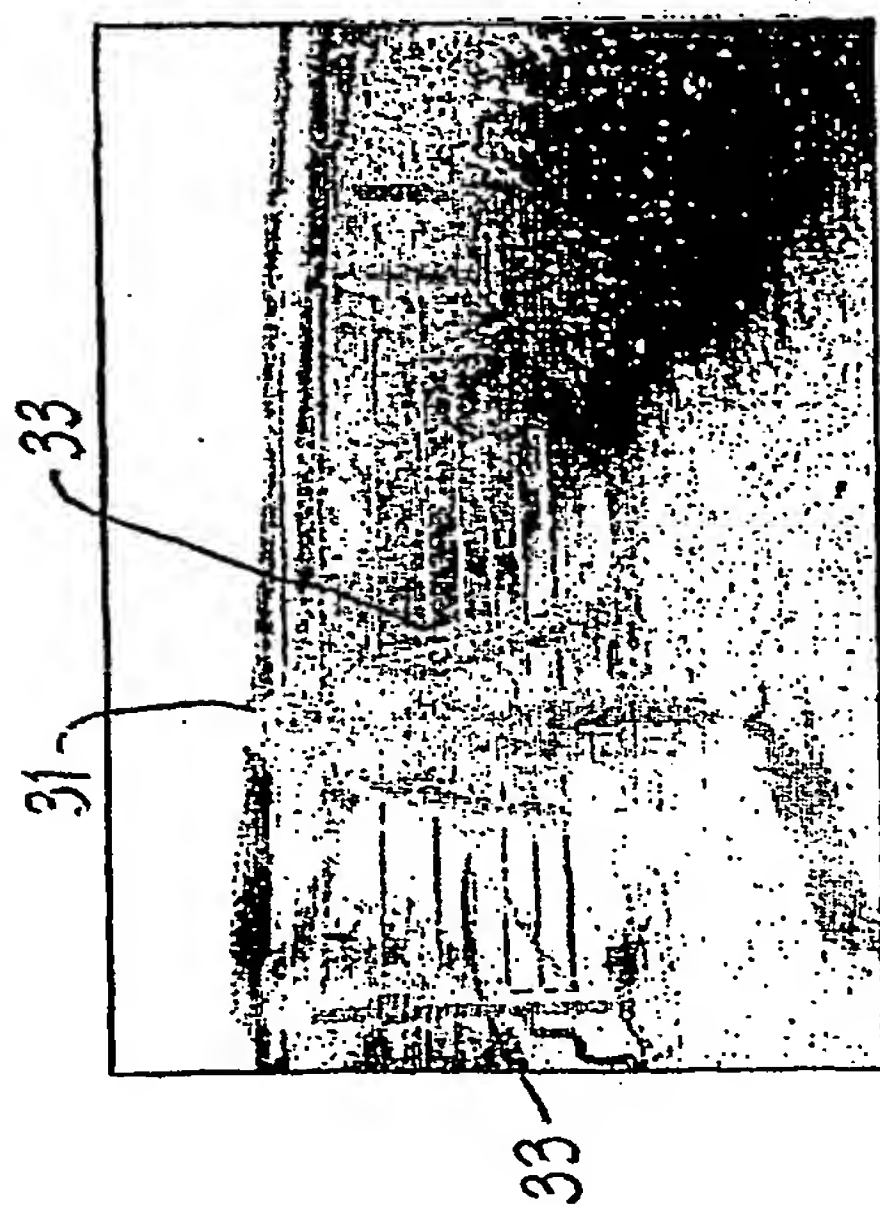


Fig. 9a

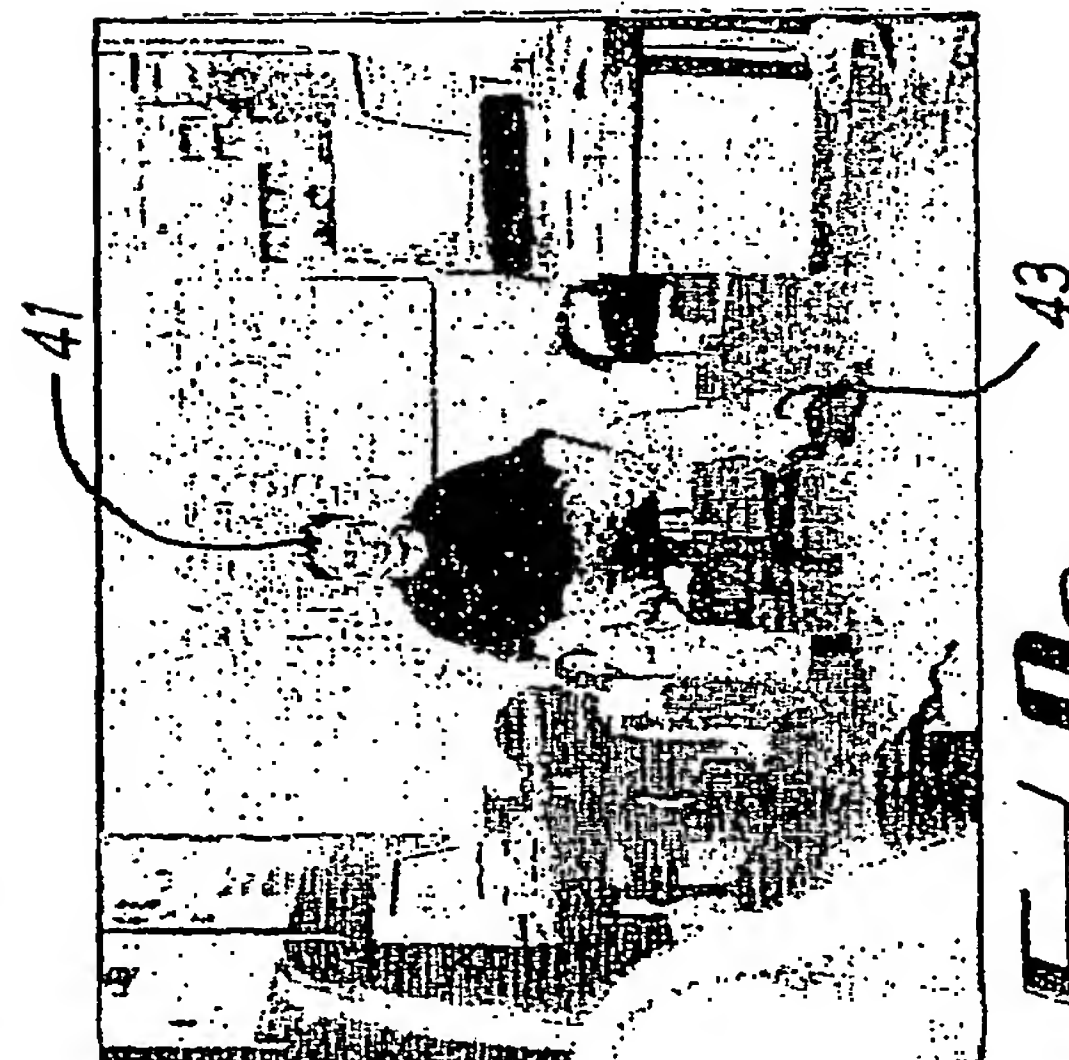


Fig. 9c

8/8

This Page is inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record

## BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURED OR ILLEGIBLE TEXT OR DRAWING
- ☒ SKEWED/SLANTED IMAGES
- ☐ COLORED OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REPERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images  
problems checked, please do not report the  
problems to the IFW Image Problem Mailbox**